October, 1982 LIDS-P-1244

APPROXIMATIONS FOR NONLINEAR FILTERING[1]

Sanjoy K. Mitter

Department of Electrical Engineering and Computer Science
and Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

This paper is based on lectures given by the author at the NATO Advanced Study Institute on Nonlinear Stochastic Systems, held at Algarve, Portugal, May 1982.

(1) This research has been supported by the Air Force Office of Scientific Research under Grant No. AF-AFOSR 82-0135.

1. INTRODUCTION

The objective of this paper is to show how recent work on nonlinear filtering can give qualitative insight into practical nonlinear filtering and suggest approximation schemes for optimal nonlinear filters.

To simplify the exposition we shall consider filtering problems in which the state and observation processes are scalar stochastic processes. We shall also present formal derivations. For the rigorous derivation of these results and precise hypotheses see the authors' papers FLEMING-MITTER [1982], MITTER [1982], OCONE [1980] and the references cited there.

2. PROBLEM FORMULATION

Consider the nonlinear filtering problem

$$dx(t) = b\big(x(t)\big)dt + \sigma\big(x(t)\big)dw(t) \tag{2.1}$$

$$dy(t) = h\big(x(t)\big)dt + d\tilde{w}(t) \tag{2.2}$$

where $x(t)$ is the state process, $y(t)$ is the observation process and $w,\tilde{w}$ are assumed to be independent standard Wiener processes. It is required to construct an estimate

$$\widehat{\phi\big(x(t)\big)} = E\left[\phi\big(x(t)\big)\Big|\mathscr{F}_t^{\,y}\right], \tag{2.3}$$

where ϕ is some suitable function. In many situations we are required to estimate $x(t)$ itself in which case $\hat{x}(t)$ is just the conditional mean. Later we shall be discussing other estimates such as the conditional mode estimate and the maximum-likelihood estimate which are based on the conditional density of $x(t)$ given $\mathscr{F}_t^{\,y}$.

3. BASIC EQUATIONS AND BASIC STRATEGRY OF SOLUTION

It is now well known and well-established that working with the conditional density of $x(t)$ given $\mathscr{F}_t^{\,y}$ need not be the right approach to the solution of the nonlinear filtering problem. Instead one works with the Zakai equation for the unnormalized conditional density $q(x,t)$ which satisfies

$$dq = A^{*}q\,dt + hq\,dy \qquad t \geq 0 \quad , \tag{3.1}$$

where A is the generator of the state process $x(t)$ and $*$ denotes formal adjoint. It can be shown that the conditional density $\rho\,(x,t)$ is given by

$$\rho(x,t) = \frac{q(x,t)}{\int_R q(x,t)\, dx} , \qquad (3.2)$$

and the estimate $\hat{\phi}\left(x(t)\right)$ given by

$$\hat{\phi}\left(x(t)\right) = \frac{\int_R \phi(x) q(x,t) dx}{\int_R q(x,t) dx} . \qquad (3.3)$$

It should be noted that the equation (3.1) is essentially a linear equation and a simpler object to analyze than the Kushner-Stratanovich equation for the unnormalized conditional density.

We shall rewrite (3.1) in Stratanovich form and write it formally as:

$$q_t = (A^* - \frac{1}{2} h^2)q + \dot{y}(t)hq, \ t \geq 0 \qquad (3.4)$$

where $\cdot$ denotes formal differentiation. Everything we say can be made rigorous, for example, by working with the pathwise filtering equation, which is obtained from $q(x,t)$ by defining $p(x,t)$ as:

$$q(x,t) = \exp\left(y(t)h(x)\right) p(x,t) , \qquad (3.5)$$

and noting that $p(x,t)$ satisfies:

$$p_t = (A^y)^* p + \tilde{V}^y p , \qquad (3.6)$$

where

$$A^y\phi = A\phi - y(t)\sigma^2(x)h_x(x)\phi_x \qquad (3.7)$$

$$\tilde{V}^y(x,t) = -\frac{1}{2} h(x)^2 - y(t)Ah(x) + \frac{1}{2} y(t)^2 h_x(x)\sigma^2(x)h_x(x). \qquad (3.8)$$

The pathwise filtering equation is an ordinary partial differential equation and not a stochastic partial differential equation. For a discussion on the pathwise filtering equations see CLARK [1978]. The basic strategy of solution schemes for filtering proceeds by analyzing equation (3.1) to answer questions such as existence of finite-dimensional statistics, invariance of (3.1) under groups of transformations and also by obtaining estimates based on equation (3.1) itself. This will be illustrated in the later sections.

4. THE DUALITY BETWEEN ESTIMATION AND CONTROL

The duality between estimation and control is understood by giving equation (3.4) a stochastic control interpretation. We follow FLEMING-MITTER [1982] in this section. To simplify the exposition let us assume $\sigma \equiv 1$.

$$q(x,t) = \exp\left(-S(x,t)\right) , \tag{4.1}$$

where we are using the fact that q is positive. This transforms equation (3.4) into the Bellmen-Hamilton-Jacobi equation

$$S_t = \frac{1}{2} S_{xx} + H(x,t,S_x) , \quad t \geq o \tag{4.2}$$

$$S(x,o) = S^{o}(x) = -\log p^{o}(x) , \tag{4.3}$$

where $p^{o}(x)$ is the initial density of $x(o)$ assumed to be positive where

$$H(x,t,S_x) = -b(x)S_x - \frac{1}{2} S_x \; S_x \quad + \frac{1}{2} h^2 - \dot{y}(t)h + b_x . \tag{4.4}$$

To get an explicit solution to (3.4) we need to solve (4.2) or equivalently the stochastic control problem

$$\begin{cases} d\xi = b\left(\xi(\tau)\right) d\tau + \underline{u}\left(\xi(\tau),\tau\right) d\tau + dw & o \leq \tau \leq t \\ \xi(o) = x , & \end{cases} \tag{4.5}$$

where the feedback control

$$u(\tau) = \underline{u}\left(\xi(\tau),\tau\right) \tag{4.6}$$

is chosen to minimize

$$J(x,t,\underline{u}) = E_x \left\{ \frac{1}{2}\int_o^t \left[\left| u(t-\tau) \right|^2 + b_x(t-\tau) + \frac{1}{2} \left| h(t-\tau) - \dot{y}(t-\tau) \right|^2 \right] d\tau + S^{o}\left(\xi(t)\right) \right\} \tag{4.7}$$

where we have added the harmless term $\frac{1}{2}(\dot{y})^2$.

For the Kalman-Bucy Filtering problem this stochastic control problem turns out to be the linear regulator problem with white Gaussian process noise, quadratic criterion, but perfectly observable. This explains the celebrated duality between filtering and control, first enunciated by Kalman.

The present formulation shows that solving the Zakai equation (3.4) corresponds to solving a nonlinear least-squares problem. Its interest lies in the fact that approximation schemes developed for solving stochastic control problems can be brought to bear on solving the Zakai equation and as we shall show in constructing estimates.

Conversely, perfectly observable, stochastic control problems can be converted by this transformation to the solution of a linear parabolic equation, and if the linear proabolic equation can be explictly solved, this would give an explicit solution to the stochastic control problem.

There are many other possibilities in using these ideas. One possibility is to factor

$$q(x,t) = \eta(x,t)\ L(x,t) \tag{4.8}$$

where $\eta(x,t)$ is a priori density for the x-process and $L(x,t)$ is a Likelihood function (unnormalized). A Bellman-Hamilton-Jacobi equation for $-\log L(x,t)$ can be obtained, which will involve the reverse Markov process corresponding to (4.5). This idea has been investigated by PARDOUX [1981] and BENSOUSSAN [1982] and can also be used to obtain estimators.

Finally, by formulating parameter identification problems as nonlinear filtering problems, we can use these ideas to treat them as nonlinear least-squares problems.

5. THE EXTENDED KALMAN-BUCY FILTER REVISITED

We now want to show how the ideas used in the previous section can be used to construct filters, and in particular gives us insight into the Extended Kalman-Bucy filter.

As an estimate for $x(t)$, one possible choice is the <u>conditional-mode estimate</u> obtained as

$$\operatorname*{Arg\,Max}_{x}\ q(x,t)\ , \tag{5.1}$$

giving rise to the trajectory $\hat{x}(t)$.

This corresponds to

$$\operatorname*{Arg\,Min}_{x}\ S(x,t)\ , \tag{5.2}$$

by virtue of (4.1).

By slight abuse of terminology, we call this the Maximum-Likelihood estimate and consider the Likelihood equation

$$S_x(x,t) = 0\ . \tag{5.3}$$

For the requisite smoothness properties which would make the sequal rigorous see FLEMING-MITTER [1982].

We obtain an equation for $\hat{x}(t)$ by considering

$$\frac{d.}{dt}\left[S_x\big((x,t)\big)\right] = 0$$

along the trajectory $\hat{x}(t)$.

This gives us an equation for $\hat{x}(t)$:

$$\frac{d\hat{x}}{dt} = b\big(\hat{x}(t)\big) + S_{xx}^{-1}\big(\hat{x}(t)\big)\Big[h_x\big(\hat{x}(t)\big)\Big(\dot{y}(t) - h\big(\hat{x}(t)\big)\Big) - \frac{1}{2}\cdot S_{xxx}\big(\hat{x}(t)\big) - b_{xx}\big(\hat{x}(t)\big)\Big]. \qquad (5.4)$$

This derivation requires that S_{xx} be invertible. It is possible to derive differential equations for S_{xx}, S_{xxx} etc. but these couple together and hence we do not get a closed-form solution.

Several remarks are in order. Firstly, in the Kalman-Bucy situation, one can easily show that S is a quadratic function and in that case the process $S_{xx}^{-1}\big(\hat{x}(t)\big)$ turns out to be independent of y and is indeed the error covariance and we recover the Kalman-Bucy filter.

Secondly, if we make the assumption that S_{xx} is invertible at $\hat{x}$, then by the Morse Lemma (MILNOR [1963]), in the neighborhood of the nondegenerate critical point $\hat{x}$, S is a quadratic in a suitable coordinate system. In this coordinate system S_{xxx} is zero and we get the structure of the extended Kaman filter, but with the additional term $-S_{xx}^{-1}\, b_{xx}$ (which is non-zero unless b is linear). Indeed, a possible choice for $S_{xx}^{-1}\big(\hat{x}(t)\big)$ is obtained by solving the Riccati equation

$$\dot{\sigma}(t) = 2b_x\big(\hat{x}(t)\big)\sigma(t) - h_x^2\big(\hat{x}(t)\big)\sigma^2(t) + 1$$

as in the Extended Kalman filter.

The invertibility of S_{xx} is connected with the observability of the nonlinear system and related to "hypoellipticity" concepts for stochastic partial differential equations and the filtering problem.

With the aid of the process $S_{xx}\big(\hat{x}(t)\big)$ one can define the analog of the Fisher-Information matrix. The whole line of enquiry has close connections to the Cramer-Rao lower bound for estimation and its generalizations to the filtering situation by

Bobrovsky and Zakai (cf. BOBROVSKY-ZAKAI [1976]), and information-theroetic ideas in nonlinear filtering (GALDOS [1975]).

Finally, the invertibility of S_{xx} is related to the conjugate point phenomenon in the Calculus of Variations.

6. FILTERING WITH SMALL PROCESS NOISE

Consider the situation where

$$dx(t) = b\left(x(t)\right) dt + \sqrt{\varepsilon}\, dw(t) \quad , \tag{6.1}$$

where $\varepsilon > 0$ is a small parameter. This case can be analyzed using the work of Fleming (FLEMING[1971]), provided a regularity condition eliminating conjugate points is imposed. In the limit as $\varepsilon \downarrow 0$, we get a Hamilton-Jacobi equation (as opposed to a Bellman equation) and we get an ordinary optimal control problem parametrized by y.

The details of this as well as a more rigorous discussion of section 5 will appear in a joint paper with Fleming.

REFERENCES

Bensoussan, A.: 1982, Unpublished Manuscript on Nonlinear Filtering.

Bobrovsky, B.Z., Zakai, M.: 1976, A Lower Bound on the Estimation Error for Certain Diffusion Processes, IEEE Trans. on Information Theory, Vol., IT-2, No. 1, pp. 45-51.

Clark, J.M.C.: 1978, The Design of Robust Approximations to the Stochastic Differential Equations of Nonlinear Filtering in Communication Systems and Random Process Theory, ed. J.K. Skwirzynski, Sithoff and Noordhoff.

Fleming, W.: 1971, Stochastic Control for Small Noise Intensities, SIAM Journal on Control.

Fleming, W.H., Mitter, S.K.: 1982, Optimal Control and Nonlinear Filtering for Nondegenerate Diffusion Processes, to appear Stochastics.

Galdos, J.: 1975, Information and Distortion in Filtering Theory, Doctoral Dissertation, M.I.T.

Milnor, J.: 1963, Morse Theory, Princeton University Press, Princeton, NJ.

Mitter, S.K.: 1982, Lectures on Nonlinear Filtering and Stochastic Control, Proceedings of CIME School on Nonlinear Filtering, and Stochastic Control, July 1981, proceedings to be published by Springer-Verlag.

Ocone, D.: 1980, Topics in Nonlinear Filter as a Likelihood Function, proceedings Decision and Control Conference, San Diego, CA.

Pardoux, E.: 1981, Nonlinear Filter as a Likelihood Function, proceedings Decision and Control Conference, San Diego, CA.